TITLE: Detecting Usability Problems with Eye Tracking in Airborne Battle Management Support

Detecting usability problems with eye tracking in airborne battle management support

Frank Ole Flemisch
Reiner Onken
Universität der Bundeswehr München
Werner-Heisenberg-Weg 39, D 85577 Neubiberg / Germany
Tel. -49-89-6004-2139 or -3452
frank.flemisch@unibw-muenchen.de reiner.onken@unibw-muenchen.de

The challenge: More information, still for humans

Information is a key element for success or failure on future battlefields. Continuous advances in information technology and battle management systems, especially growing computer capacity and interoperability promise to provide comprehensive tactical situation awareness down to unit level, thereby improving mobility, survivability and sustainability of today's weapon systems.
However increased availability of information in the computerized support systems does not automatically lead to increased usability. It rather may lead to information proliferation, hidden information and pertinent problems regarding operator information processing. These problems even grow under time pressure in a stressful environment. Are these problems unavoidable? Or is there a solution to handle the overwhelming amount of information which tomorrows battle management systems and personal have to work on?

In aviation there were tremendous technological efforts during the last twenty years to answer similar questions through increase of automation like the introduction of flight management systems or fully computerized „glass cockpits". Again, to the surprise of many people, the relative safety did not increase, but remained almost constant [Billings 1997].

The upcoming solution: Cognitive Automation and Assistant Systems

These problems led to discussions and doubts about the benefit of automation on the one hand, and research in favor of "cognitive automation" on the other hand. As opposed to increased conventional automation in the sense as mentioned above, **cognitive automation** is based on cognitive engineering (e.g. [Rasmussen et.al. 1994]) and more adapted to interact with human cognition [Onken 1998]. This gives the chance to handle more information in the cockpit without decreasing usability.

Prototypes of cognitive automation in aviation are the Cockpit ASsistant SYstem CASSY for civil IFR, flight tested in 1994, and CAMA, the Crew Assistant Military Aircraft, developed together with DASA, DLR , ESG and the University of Armed Forces. Simulator trials were conducted in 1998, flight tests are scheduled for 2000, e.g. [Lenz & Onken 2000] in this proceeding.

But: How can we be sure that no new problems will arise with cognitive automation?

Undoubted, conventional automation was motivated by positive intentions. One major intent was the reduction of workload. The effect was so enormous that, as a result, we face now a "pilot-out-of -the-loop" problem, e.g. [Endsley &Kiris 1995], the "ironies of automation" [Bainbridge 1987] and operators speaking of "99% boring, 1% panic" [Kraiss 1994].

How can we be sure that cognitive automation solves problems but does not raise new problems? If we can not be sure, how can we learn from the lessons and implement ergonomics / human factors right from the start of the development cycle?

Ergonomics / human factors offer a wide range of methods for detection and handling of usability problems. On the other hand, even well experienced concepts like e.g. workload more often fail to reliably describe the problems, especially with increasing technical complexity or „self animated" machines [Sarter & Woods 1994]. How can we implement newer concepts like usability [Nielsen 1993] or situation awareness [Endsley 1995], how can we detect problems like cognitive fixation or dangerous attention distribution?
How can we meet the often different demands of our target groups such as engineers, managers, scientists and operators?
How can we bridge the gap between the diametrical poles "subjective / objective", "intuitive / analytical", "global / detailed" or "scientifically exact / efficient" in order not only to detect but to solve usability problems?

Paper presented at the RTO HFM Symposium on "Usability of Information in Battle Management Operations", held in Oslo, Norway, 10-13 April 2000, and published in RTO MP-57.

A prototype for integrated usability testing: caSBAro

As an answer to these needs a new kind of usability testing tool, caSBAro, was developed in parallel with CAMA. The acronym stands for:

c omputer	supports not replaces human
a ided	factors analysis
S ituation and	analysis of behavior cannot
B ehavior	be done without analyzing the
A nalysis	underlying situation
r eplay and	the record can be fully replayed in a flight simulator
o nline	all caSBAro analysis modules must be capable to work in realtime for the future option to plug them into the assistant system

Figure 1 shows the structure of caSBAro: a generic flight simulator, eye- and headtracker, digital videodisc system and recording / visualization / analysis of man- and machine behavior.

One core element of caSBAro is the sharpening of our best usability measuring tools, our pilots, by offering them a full mission replay in the simulator including the eye tracking records. This gives engineers, managers and operators the platform for a very detailed debriefing without memory effects, an intuitive access to objective data analyzable down to the byte and eyeblink level [Flemisch & Onken 1999].
Another core element of caSBAro and focus of this paper is the analysis of the operators interaction resources, especially the distribution of the visual resource in the cockpit. This gives an almost direct access to the visual part of the human bottleneck and usability problems like information overload or dangerous attention distribution.

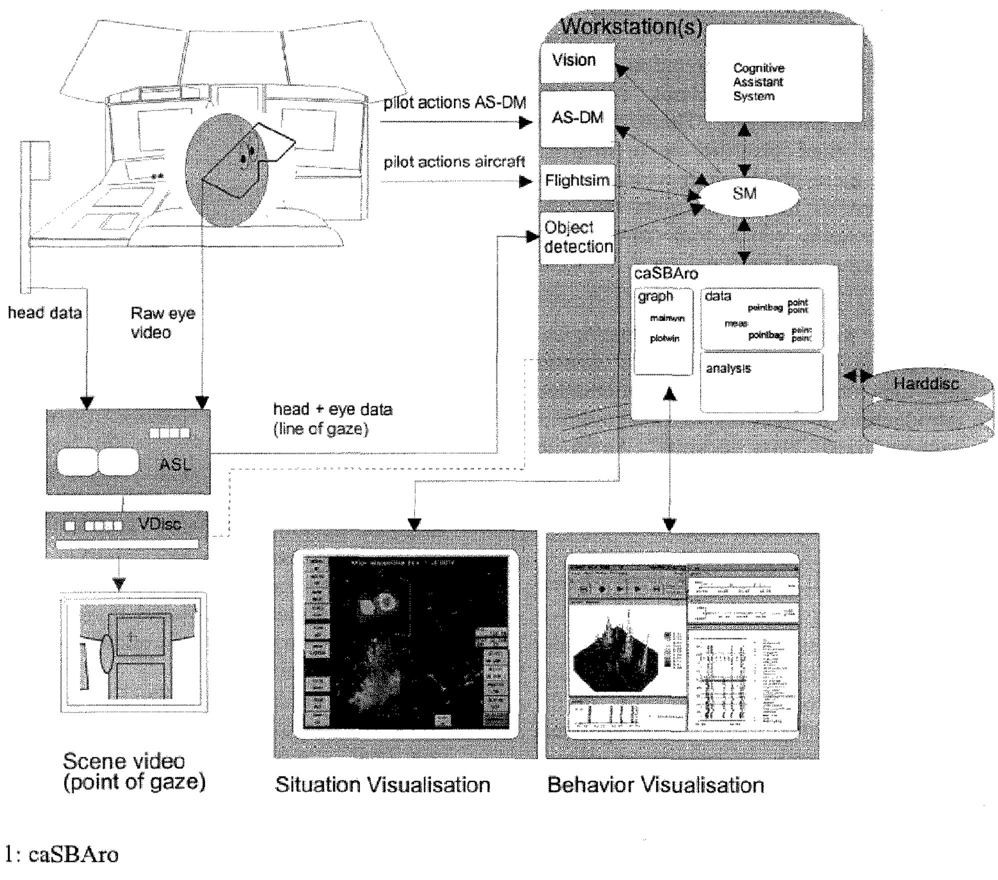

Figure 1: caSBAro

Experimental series on "variation of technical support for manual flying and navigation"

The main objectives for the following series of generic simulator experiments (or better: quasi experiments in the rigorous sense of classical experimental psychology) were:

- estimation of the method's overall sensivity for the visual resource,
- estimation of the method's potential in the ergonomical toolbox, compared to the classical methods "subjective workload" and "objective performance",
- exploration of relationship between different technical supports and their effects on the operator's visual resource in order to improve the assistant system CAMA.

The subjects were 6 military pilots of a German Tactical Air Transport Wing (LTG61 Landsberg), aged 30 –41 (average34) with a experience of 800 – 6000 (average 2700) flight hours on several aircraft types, especially the two engine transport aircraft C-160 "Transall". The experiments were embedded into a 2 days / pilot simulation campaign.

The task performed by the pilots consisted of a combination of two subtasks, a tracking subtask with higher frequency (manual low level flight of a preplanned minimum risk route), and a low frequency supervision / navigation subtask. Each subtask was supported by different technical means.

On the one hand this prototype combination of subtasks is quite relevant for the aviation domain, on the other hand it promised to be prototypical enough to allow a transfer of experience into other domains.

The scenario consisted of a preplanned low level minimum risk route with about 7min flight time in a hilly area (Black Forest), a dynamic threat theater with simulated hostile SAM-stations (Surface-to-Air Missiles) and an ACO (Airspace Control Order) with egress corridors.

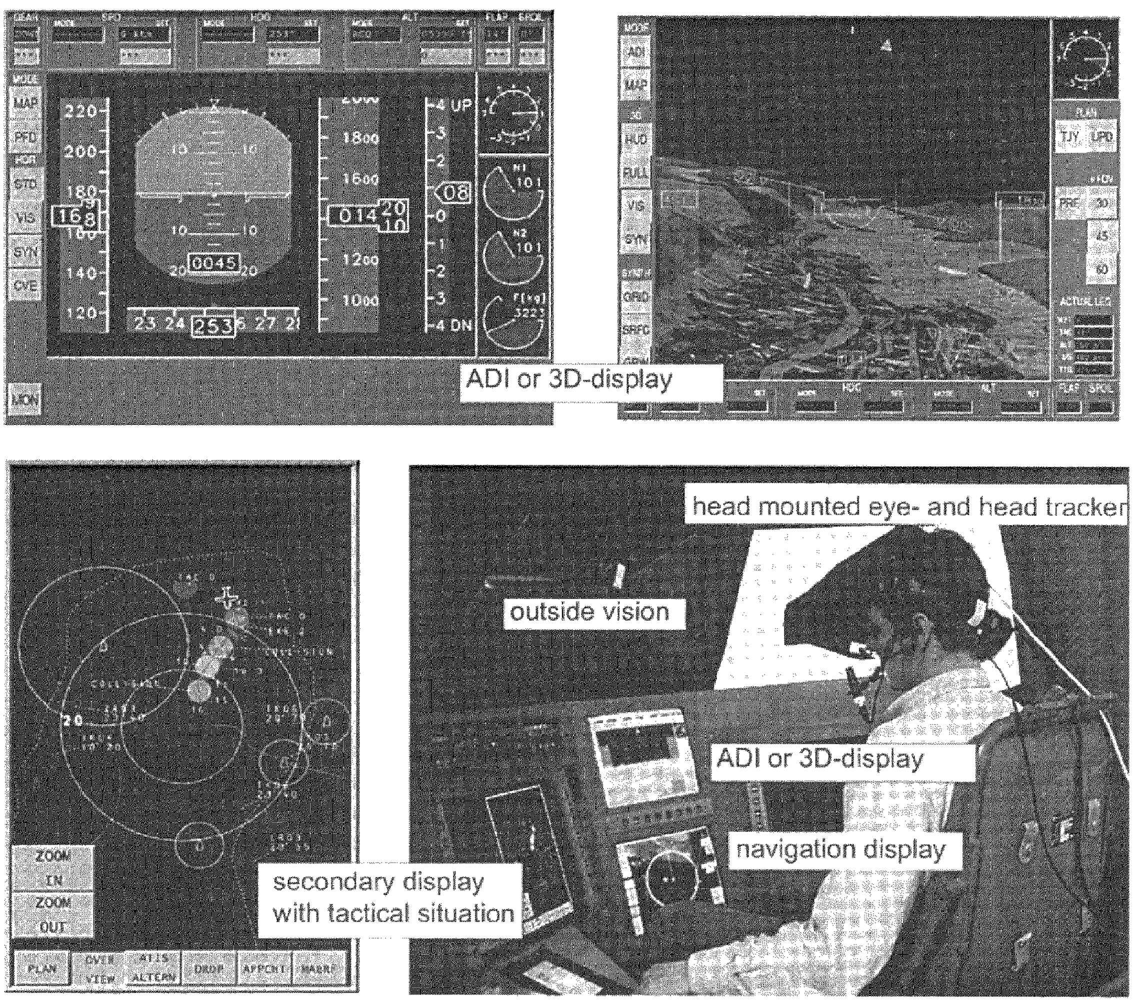

Figure 2: simulator with displays and eye tracking equipement

The subtask F "Manual Flight" demanded flying of the minimum risk route, which remained constant through all experiments, under VFR-conditions (Visual Flight Rules) and "as accurately, fast and, most important, safely as possible". The technical support for this subtask was varied as follows:
ADI: Classical combination of cockpit instruments with artificial horizon, speed, altitude, radar altitude etc. as in state of the art civil aircraft glass cockpit aircraft (as shown in figure 2).

3D: newly developed flight guidance display with three-dimensional dynamic picture of terrain elevation, terrain features and a "tunnel in the sky" of the minimum risk route (as shown in figure 2), (by ESG, see also [Schulte & Stütz 1998]).

3DADI: newly developed combination of ADI and 3D, much smaller 3D-display area.

Auto: no manual but automated following of the minimum risk route, [Bamberger & Lenz 1998].

The subtask N "Navigation" consisted of

1. Monitoring of the tactical situation on the secondary display with regard to changes.
2. In case of changes: decision whether own route or egress corridor is endangered, callout ("threat factor / no factor").
3. If route is endangered: choice of alternative egress corridor, callout of choice, and
4. finally replanning by selecting the alternative corridor on the Secondary display (touchscreen), then selecting button "Replan via" on the Navigational Display.

After that the replanning sequence was terminated, the original minimum risk route remained constant during all experiments.

The technical support for this subtask was varied as follows:

No support: "only" visualization of the tactical situation on the Secondary Display.

Highlighting: highlighting of changes by different color and blinking symbols.

Callout: (in addition to highlighting) a speech output "tactical situation changed".

Proposal: (in addition to highlighting and callout) a machine generated solution by speech output, .e.g. "replan via corridor TK05", highlighting of the alternate corridor and textual feedback on the navigation display.

Simplified activation: (in addition to all support mentioned above) the simple activation of a proposal by selecting a "Roger Do It" button or alternatively by a speech input "roger do it".

Variation and combination of subtasks and support:

Comparison I (E_1 – E_4) investigates subtask F "manual flight" with different technical support, but with no navigation (See also table 1).

Comparison II (E_5 – E_9) addresses the navigational subtask N with different technical support, combined with a pseudo flight task "supervision of automated low level flight". These conditions are comparable to those of a PNF (Pilot Non Flying) busy with a navigational task.

Comparison III (E_15, E_12, E_14, E_11) deals with the combinations of the two subtasks with none or complete support. The idea was that extreme combination of support would also generate extreme behavior and would therefore stretch out the behavioral spectrum in a manner that in between, nonextreme combinations can be derived at least qualitatively by interpolation of extreme combinations without being measured explicitly. The simple but striking reason behind this is the limited maximum time for the experiments due to the weight and pertinent discomfort of the head mounted equipment.

To minimize effects of order of the test runs, they were not conducted in logical order, but were, after placing the order-critical experiments (see chapter collision aircraft), varied according to a replicated Latin square design (see also [Johannsen & Rouse 1983]).

			subtask F „manual flight“			
			autopilot	flight guidance display		
				ADI	3D	3D-ADI
subtask N „Navigation“	no Nav.	**Comparison I →**	E_4	E_1	E_2	E_3
	support	display only	E_5	E_15	E_12	
		+ highlighting	E_6			
		+ callout	E_7	**Comparison III**		
		+ proposal	E_8			
		+ simplified activation	E_9	E_11	E_14	
			Comparison II			

Table 1: Variation of technical support, combination of subtasks

Eye tracking data

Figures 3 – 5 represent the distribution of the visual resource across the visual workspace for the specific subtask / support combination, averaged over all pilots and flighttime. The lighter the areas are, the more fixation time (in this case corresponding to visual attention) pilots spent on that particular spot (excluding warm up phase, exponentially accumulated fixation time, shifted to positive values, standardized to volume integral and projected into 2D, graphical representation by caSBAro-XRT, [Morawski 1999]).
The white %-numbers represent the average percentage of visual attention on the specific region of interest (displays, outside vision)

Subjective workload with SWAT rating

In order to allow comparison of eye tracking with classical approaches, the subjective mental workload of the pilots was measured with the SWAT method (Subjective Workload Asessment Technique). According to this method, mental workload contains three components, time pressure T, mental effort E and stress S in three stages, low 1, medium 2 and high 3.
The TES-triple in figures 3 –5 represent the pilots' median postflight estimation of subjective workload.
W represents the mean value of the conjoint subjective workload. This "conjoint scaling" method also takes into account interpersonal differences in the relative importance of T, E and S. Part of this method is that pilots sort the 27 possible SWAT-combinations in order of relevance before the experiments [Nygren 1991].

Performance P_F for subtask F "Manual Flight"

As the above mentioned subjective workload is only sensitive for the overall task combination, the relationship between technical support and specific subtask must be evaluated by subtask sensitive methods. Subjective methods, e.g. Cooper-Harper-Scale, would also be usable here, but because of the caSBAro capability for recording aircraft parameters, the calculation of a "mean distance to a specified track" d_m as most frequently used method for objective performance assessment can easily be done.
Mean speed ias_m helps to detect potential speed accuracy tradeoffs.

Performance P_N for subtask N "Navigation"

Like for P_F, speed accuracy tradeoffs also have to be controlled for P_N. This is done by two values representing time and accuracy: overall time for solving a conflict and percentage of correct / successful reaction.

Moreover this subtask can be structured with respect to the different stages of human information processing, e.g. according to [Wickens 1992]:

1. **per**ception, here detection (and callout) of a potential conflict (step 1 and 2 of the description for subtask N above).
2. decision and response **sel**ection, here selection of a alternative egress corridor (and callout).
3. response **execution**, here activation of a replanning process.

Because a specific technical support can have different effects on different stages, average time and quality percentage was calculated for each specific stage. In order to highlight the overall effect, only correct reaction were accumulated over the three stages. Table 2 provides an example referring to figure 4 E_5:

	Single stage performance	Accumulated overall performance	time
perception		82%	3.7s
selection	88% + 2.0s	72%	5.7s
execution	60% + 3.6s	**43%**	**9.3s**

Table 2: objective performance of subtask N, example from figure 4, E_5

This means that within this subtask/support combination, averaged over 6 pilots, 82% of all navigational conflicts (changes of tactical situation that endangered the preplanned route) were detected and called out by the pilots after 3.7 seconds. 2 seconds later 88% of these conflicts were also solved (and the solution called out) correctly, 3.6 seconds later 60% of these solutions were also executed correctly, so that 43 % of all conflicts were solved correctly after 9.3 seconds, 57% were incompletely replaced by a subsequent conflict or failed at one or the other stage of the pilot's information processing.

Comparison I: Variation of subtask F "Manual Flight", no Navigation

Comparison I looks at the isolated subtask F "Manual Flight" with different flight guidance support ADI, 3D, 3DADI and automatic flight. In figure 3 e.g. "P_F+" stands for an improvement in flight performance, "W =" for an almost constant subjective workload. Black arrows show a virtual flow of visual attention between two configurations.

E_1 ADI represents the classical low level flight under VFR conditions (Visual Flight Rules) and with state of the art displays: Subjective conjoint workload W is average with 42%. This subtask and configuration is the daily but nevertheless not easy job of these pilots. Visual attention is mostly (56%) directed to the outside vision, where e.g. hill ridges are fixated in order to avoid terrain collisions. The visual scanning pattern of the ADI is characterized by a classical "basic T", a repetitive change between speed, artificial horizon and altitude / radar altitude / variometer. Short gazes downwards to the Navigational Display are used to detect deviations from the minimum risk route and to perform medium-term orientation ("ok, after the next ridge right into the valley, then one mile straight on, uups...").

E_2 3D is the same flight with 3D-display: Visual attention is attracted by the integrated information of terrain, aircraft attitude and minimum risk route on the 3D-display. This limited visual resource is withdrawn mostly from the outside vision and partly from the navigational display. Some pilots urge themselves to check the outside environment more frequently (max. 35%), others just abandon this source of information (min. 4%). Flight path accuracy as measurement of objective performance is almost 4 times higher than with classical ADI, speed is higher, subjective workload is clearly reduced.

E_3 3DADI is the hybrid of classical ADI and 3D: The concentration effect already observed in E_2 even grows stronger, performance is almost equal, subjective workload is increased due to the small size of the 3D-window, but is still lower than E_1.

E_4 autopilot with pilot as supervisor: Even though the autopilot configuration is quite convenient (lower flight path accuracy and speed as flown by the pilots themselves), subjective workload is higher than in e.g. E_2. When asked about these surprising ratings pilots stated a "natural distrust" of automated flight due to lack of experience and short reaction time in case of malfunction.

The automation frees visual resources, which flow into the secondary and the navigational display, nevertheless the overall distribution of visual attention is quite similar to E_2. As e.g. the scanpath theory [Stark & Choi 1996] formulates a strong relationship between observed visual behavior and internal mental representation of a visual task, we can therefore assume that the visual parts of "flying an aircraft" and "supervising a machine flying an aircraft human-like " have quite similar mental representations. This affirms e.g. efforts like [Schulte 1996], who investigated visual behavior of pilots in low level flight by stimulating them with a movielike video replay of a real flight in a simulator with outside view.

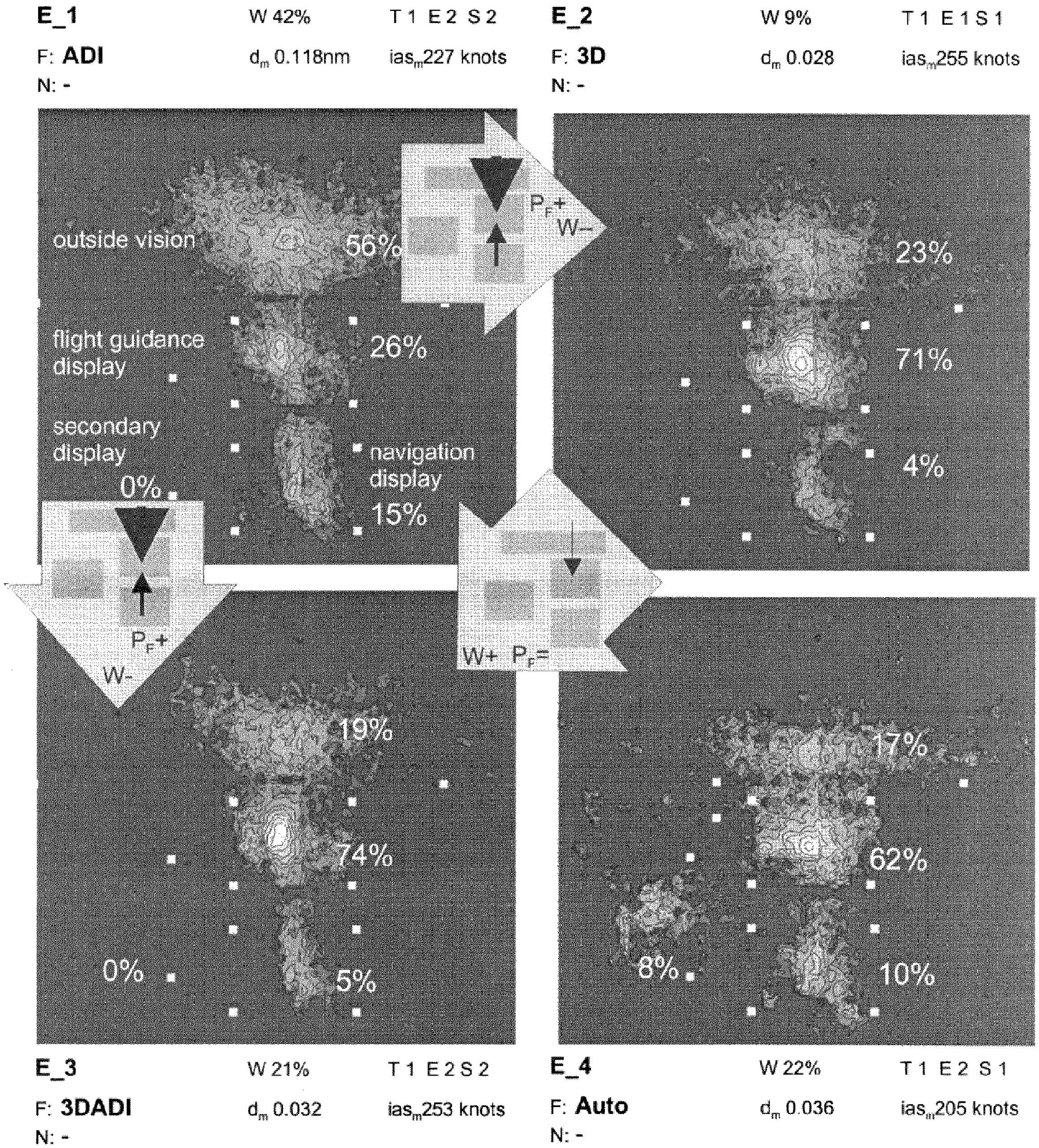

Figure 3: Comparison I, flight with different displays, no navigation

Comparison II: Variation of Navigation, Autopilot

Comparison II (figure 4) investigates the influence of different technical support for subtask N "Navigation" without subtask F.

E_5 without support: 89% of the visual attention is located at the secondary display, only sometimes gazes are moving elsewhere, e.g. to the outside view. Subjective workload is lower than e.g. E_4 (supervision of autopilot), objective performance is only medium mainly because of execution interruptions by new conflicts. This gives evidence that the experiment is working close to the upper limits of performance and is therefor sensitive.

E_6 with highlighting: rate and speed of detection increases. Reasons for that might be a better detectability in peripheral vision and a faster discrimination between endangering and harmless tactical elements. The values for selection and response execution together with pilots' comments could be a hint that the improvement is partially compensated by distracting effects caused by the symbol blinking.

E_7 with additional speech output in case of a tactical change: performance and subjective workload are almost unchanged compared to E_6, but a fundamental quantitative and qualitative change of the visual behavior can be observed: Free visual resources almost doubled. The attentional field, which was almost exclusively focused to the navigational task / secondary display, is partially freed now. In contrast to E_6, the complete right side of outside vision can be covered now.

E_8 with additional proposal for conflict resolution: high improvement of response selection, slight reflux of visual attention into the secondary display. However, regarding the overall performance an almost paradox effect can be observed: Although pilots know the conflict solution much faster than the machine, they tend to wait for the proposal to assure themselves. So they loose precious time for the execution before the next conflict occurs. This effect could of course also happen in reality, but the observed effects on the overall performance can be considered as an artifact caused by the experimental conditions, especially the relative simplicity of the navigational task.

E_9 with simplified activation by "roger do it" button or speech input: The "waiting for the proposal" effect is still observable, but these proposals are activated fast and accurate, so that compared to unsupported E_5 overall time is equal, but quality doubles! Freed visual resources can flow in other information sources.

Comparison III: Extreme combination for flying and navigation

Comparison III (figure 5) investigates the extreme combinations, ADI or 3D for manual flying subtask, no support or full support including proposal and simplified activation for the navigation subtask.

E_15 - flying with ADI, navigation with no support is - not surprising - the experiment with the highest subjective workload. The flying subtask is, compared to E_1 with no navigation, performed without major dropouts, even with 20% of the visual resource withdrawn from this subtask and used for the navigational subtask. Obviously this is not enough to perform this subtask sufficiently, leads to the lowest success of 12% and a SWAT stress value of 3 for all pilots. Remarkable is the still successful rule of prioritization "aviate – navigate – communicate – manage systems"

A closer look at the extreme transfer from E_15 to full support E_14 (diagonal arrow in the center of figure 5) shows a dramatic reduction of subjective workload and a huge improvement of the subtasks' performance, especially for the navigation subtask. Regarding the visual resources, the percentage of the three information sources navigation display, secondary display and outside vision is reduced to a half and focused to the 3D display (triplication). The detailed mechanism of this resource flow becomes transparent by a closer look to the intermediate combinations:

The transfer from E_15 to E_12, ADI to 3D with unsupported navigation, leads to an improvement of flight performance with a concentration of visual resources, flowing from outside vision and navigation display into the 3D display, an effect that can also be seen in Comparison I. Better support for the subtask F does not only improve flying, moreover freed resources can be used for the navigation subtask, visible in a higher percentage of the visual resource allocation in the secondary display and a better performance on all stages of information processing.

Adding the navigation support (E_14) now leads to an acceptable performance of navigation with almost constant flight path accuracy. Simultaneously freed visual resources can be reinvested into the outside vision.
A similar picture can be developed by following the circle counterclockwise from E_15 via E_11 to E_14: additional navigation support in E_11 leads to a higher navigation performance, which reaches not yet the maximum of E_14.

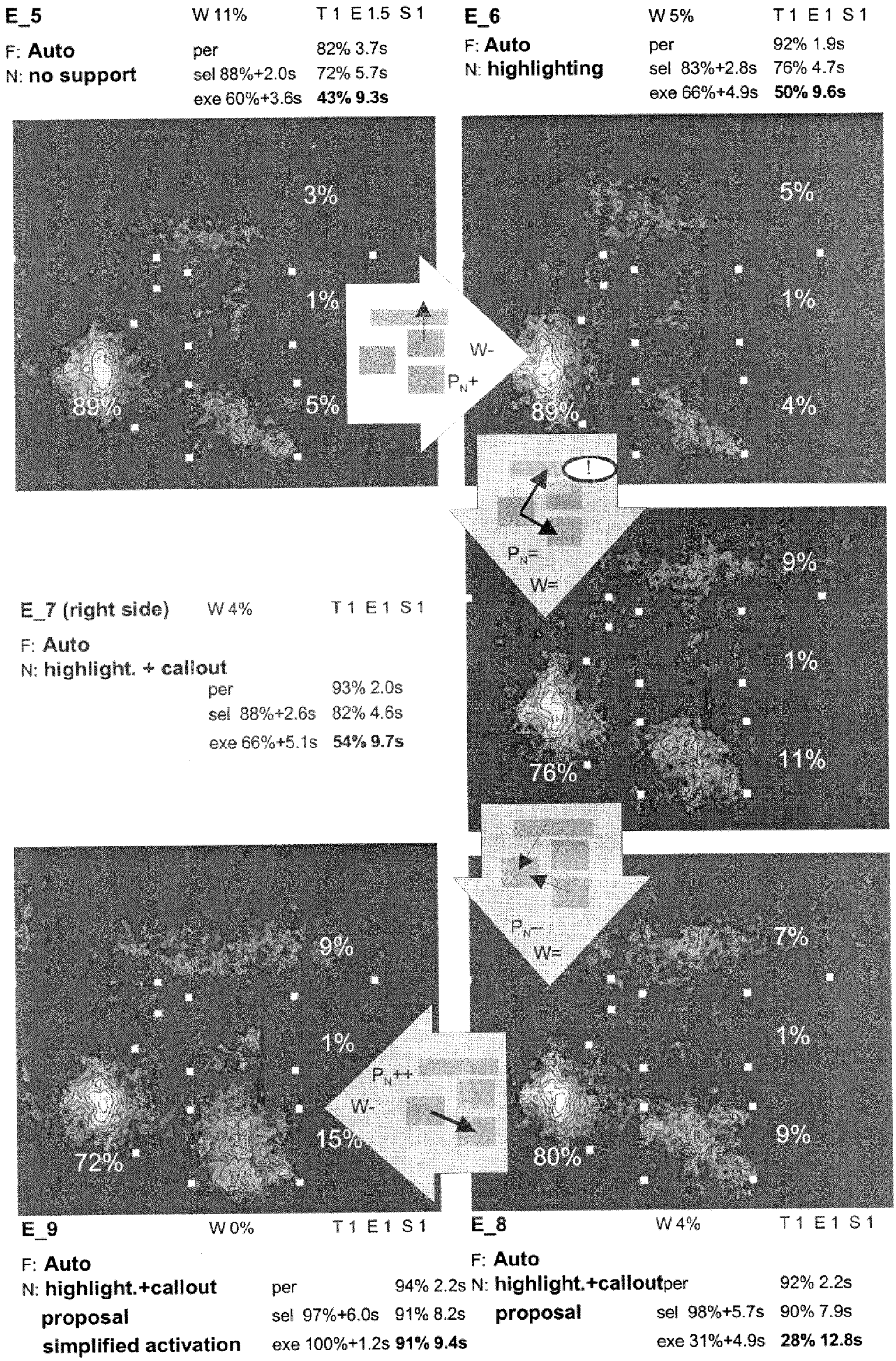

Figure 4: Comparison II, automatic flight with navigational support

Simultaneously freed resources flow back to the subtask F. These resources are reinvested not so much into the outside vision – obviously this percentage is already high enough compared to e.g. E_12 – but more into the ADI.

The transfer from E_11 to E_14 once again shows the effects of the 3D display, improvement of flight quality and concentration of visual resources.

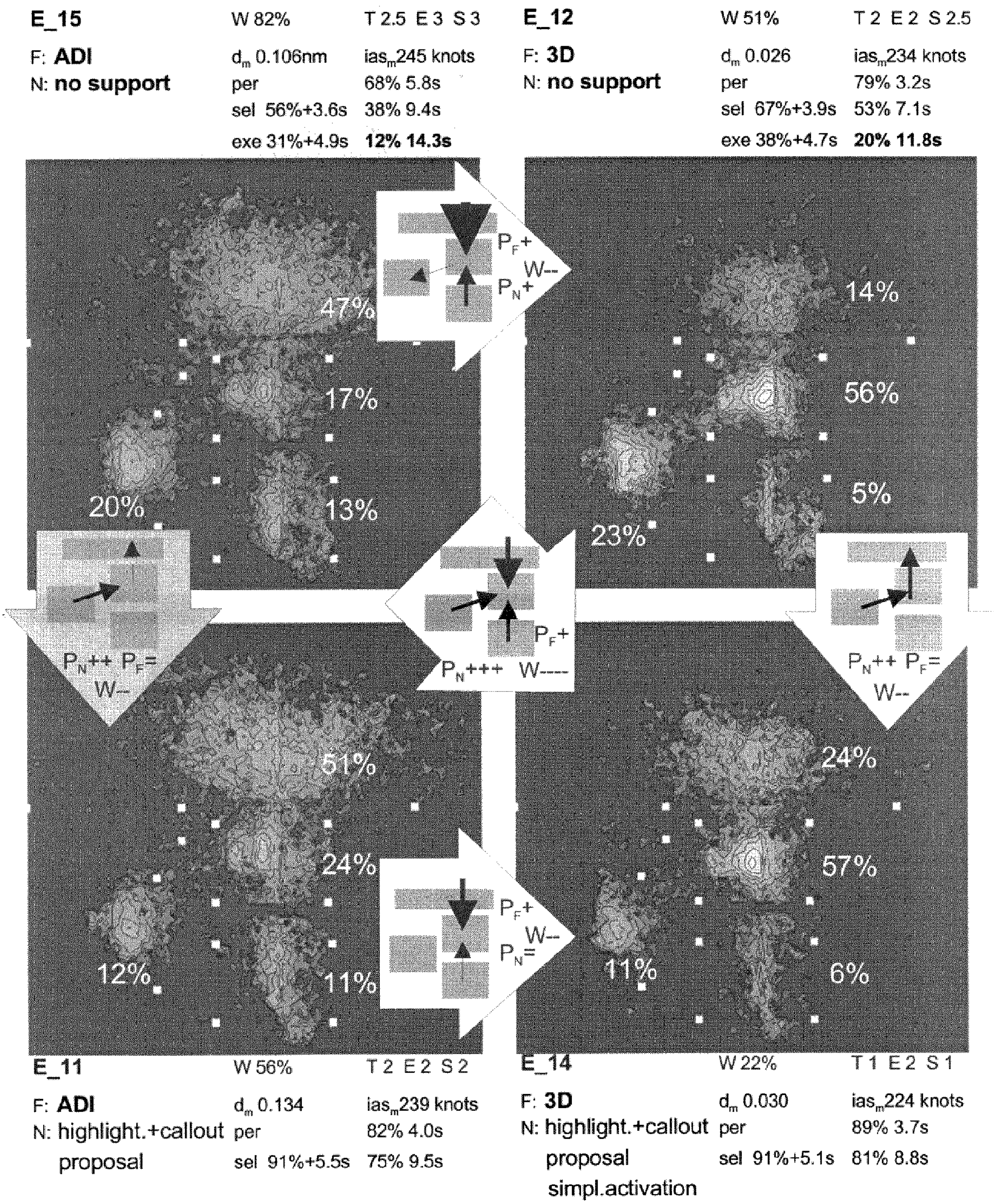

Figure 5: Comparison III, extreme combinations

Test with collision aircraft

The observed concentration of visual resources into the 3D display is not negative by itself, the performance improvements are quite impressing. Nevertheless pilots and evaluators had uneasy feelings after a look at the eye tracking videos and resource distribution. The reason is the - until now unmentioned nevertheless vital - subtask of airspace observation. This subtask has always be performed when flying according to VFR (Visual Flight Rules) in order to avoid collisions with other aircraft.

As missing visual attention is a strong indicator for missing situation awareness, and missing situation awareness is a strong contributing factor for accidents, these distributions for visual attention measured here would be enough reason to take corrective actions. Nevertheless we did an explicit test in the simulator by introducing collision aircraft. They flew along the same minimum risk route just into the opposite direction, with a speed of 200 knots, clearly visible in the outside vision.

According to signal detection theory, e.g. [Wickens 1992], after the detection of the first collision aircraft there would be a strong risk of a complete change of the attention distribution. Therefore this event "detection of collision aircraft" should reasonably happen only once per pilot without giving any hint before.

Because of the statistical difficult low number of test subjects, the pilots were asymmetrically divided into two subgroups, subgroup I with ADI / no navigational support (4 pilots), subgroup II with 3D / no navigational support (2 pilots). At the end of the corresponding flights E_15 and E_12 three successive collision aircraft were simulated. After the first detection, ascertained by callout, avoidance maneuvers or clear hints in the eye movement monitor, the experiment was terminated. The events "aircraft detected" and "aircraft not detected" had the following distribution:

Subgroup	Aircraft detected	Aircraft not detected
I ADI	4	1
II 3D	1	5

The basic hypothesis H_0 states that the two different technical configurations do not produce a different risk of colliding with another aircraft. A Pearson-$\chi 2$ test shows a significant difference with $p_{alpha} = 0.036$, but because the actuarial expectation value per cell of the 4field table is smaller than 5, it is not appropriate in this case. Luckily the side sums are almost equal and due to the experimental design a binomial distribution can be assumed, therefor the "single sided Fisher Yates exact test" can be used. This value, $p_{alpha} = 0.067$, is not significant at the confidence level used for scientific experiments (95%), but due to the lower demands of the usability paradigm, e.g. an appropriate confidence level of 90% suggested by [Nielsen 1993], H_0 can be rejected with "strong tendency to significance".

The direct transfer of this result from the small number to a complete population of pilots is, due to the design of the experiment, still not statistically valid without further control. Theoretically this outcome might have been produced by a completely different visual behavior of the 2 "collision" pilots compared to the 4 "normal" pilots and the total population. But as in E_12, the average percentage of visual attention in the outside vision is 14% for all 6 pilots, compared to a slightly smaller 12.5 % for the two "collision" pilots, there are strong hints that the danger of not detecting collision aircraft is not caused by interpersonal differences but by the configuration of displays.

Discussion of the technical support

Due to the small number of subjects the above mentioned observations and results just have tendency to significance ($p_{alpha} < 0.1$) and therefor - according to classical experimental psychology – want to be used with caution. Considering the lower statistical demands of the usability paradigm, e.g. in [Nielsen 1993], and the early phase of the exploratory process, we can nevertheless discuss the following findings:

Each of the described levels of support for the navigation subtask improves speed and/or quality of performance.
Intelligent highlighting using the situational knowledge of the assistant system improves information perception. Additional acoustic information can solve captivation of the attentional field and therefor avoid blind areas, as E_7 (Comparison II) shows. Negative effects of cluttering other acoustic information sources, which were not investigated here but can be suspected, can probably be avoided by nonvocal, spatial coding of the acoustic signal.

The machine generated proposal for conflict resolution, which was investigated here, is relatively simple due needs to keep the experiment under control. In situation with low workload pilots solve these conflicts much faster. But even with that simplicity, in situations with higher workload, especially with an additional higher frequency subtask which competes for concurrent resources, a computer-generated proposal clearly improves speed and quality of conflict resolution. It is of course mandatory, beside high quality and reliability, that the computer solution is plausible and transparent in order to build up appropriate trust / mistrust and therefor enable successful supervisory control.

The simplified activation of proposals offers an additional speed and quality improvement, which can be used optionally: In situation with sufficient resources, pilots can choose a different, more explicit man machine communication in order to maintain situation and process awareness, in situations with lack of free resources pilots can activate very simply and reliably a solution that is, at least, safe. We call that optional aspect "implicit support of operators' own resource adaptation" or "implicit adaptation". The machine does not explicitly adapt to a low resource situation, but offers implicit means for resource adaptation (see also [McKinley 1985], [Verwey 1990]). Few negative effects like potential risk homeostasis and complacency have been observed. They have to be compensated by e.g. supervised training (e.g. with mission-replay in the simulator).

The 3D display with an information fusion for terrain, flightpath and aircraft's attitude offers benefits, but there can be a problem with the concentration of visual resources toward the head-down displays. This effect, in these experiments, led to a clear lack of situation awareness regarding collision aircraft. The above mentioned simulator test investigates – of course - the configuration without navigation support, which promised to be most sensible for this effect. An influence of the head mounted equipment can not be excluded, the pilots might have been conditioned to a simulator environment where there was no experience with collision aircraft. Moreover this concentration effect will be of quite different impact with a two or three man crew.
Nevertheless it must be assured that the existing risk will be compensated. Only if this proves successful, the observed clear improvement of flight performance can fully exploited. The freed resources can be used to improve other subtasks like navigation, an effect which will be even stronger in degraded visual conditions, which where not investigated, so far.

Discussion: Is eye tracking worthwhile?

Eye movement measurement offers deep insights into man machine interaction and the mental processes of pilots. The analysis of the visual distribution in the cockpit, averaged over pilots and time, illuminates global effects of the visual resource with high qualitative depth and face validity.

Visual attention is a limited resource and has to be scheduled by the pilots to different information sources. Technical means influence this operator's own resource management positively or negatively even to the extreme of total cognitive fixation to one technical subsystem. A direct relationship between the risk of low performance, which can often not directly be measured, and an unfavorable visual distribution, which can be measured, clearly exists and can be used to detect resource based usability problems and avoid fatal results.

But these experiments also show that the methods used are not equally sensitive and reliable for all ergonomical questions. There are quite some examples in the described experiments where only one method succeeded in detecting a specific fact while the others were insensitive. A holistic qualitative picture of a specific man machine interaction seems to get illuminated best with an appropriate combination of methods.

Therefor the analysis of the visual resource is just one additional, but powerful tool in the tool box of ergonomy. Factors like time, personal effort and money will contribute to the decision whether this tool will be used. The ongoing development of smaller and cheaper hardware, the availability of sophisticated analysis software and a caSBAro like high integration of eyetracking into the usability laboratory will make it easier to use this method in the development process.

Conclusion

The benefits of information technology ought to be exploited also for battle management operations, but we know that there might be side effects and new risks like violations of the human limitations of cognition and information processing.

There are methods to control these risks, we have to use these methods right from the beginning of a development process, and we have to improve these methods permanently in order to catch up with the speed of technology.

Even if these methods are no guarantee for ideal information systems, they offer a much better chance for improving usability. If we do not take this chance, we will spend money on new technology, but will loose systems and men instead.

Acknowledgement: This paper would not have been possible without the work of flight captain and Dipl.-Ing. Marius Morawski. On July 17th 1999, two months after finishing his master thesis and 37 years old, he had a heart attack and died, unexpectedly and beyond our comprehension.

Literature

[Bainbridege1987] Bainbridge, L.: Ironies of Automation; In: New Technology and Human Error; Ed. Rasmussen, Duncan, Leplat; Wiley, 1987

[Billings 1997] Billings, C. E.: The Search for a Human-Centered Approach; Lawrence Erlbaum, New Jersey 1997

[Endsley & Kiris 1995] Endley, M.; Kiris, E.: The Out-of-the-Loop Performance Problem and Level of Control in Automation; Human Factors, 37(2), 381-394; 1995

[Flemisch & Onken 1999] Flemisch, F.O.; Onken, R.: caSBAro, a Computer Aided ErgonomicS tool to detect and solve problems in human-computer interaction; *International Conference on Computer-Aided Ergonomics and Safety CAES*, Barcelona, Spain, 1999

[Johannsen Rouse1983] Johannsen, G., Rouse, B.W.: Studies of Planning Behavior of Aircraft Pilots in Normal, Abnormal, and Emergency Situations; IEEE Transactions on Systems, Man and Cybernetics, Vol. SMC-13, No3, May / June 1993

[Kraiss 1998] „99% Langeweile und 1% panische Angst“ –über die Schwierigkeiten beim Umgang mit hochautomatisierten Systemen; In: M. Kerner (Hrsg.) Technik und Angst. Zur Zukunft der industriellen Zivilisation. Aachen, 1994

[Lenz & Onken 2000] Lenz, A.; Onken, R.: Crew Assistant Military Aircraft CAMA; RTO HFM Symposium Oslo 2000

[McKinley 1985] McKinley, W.H. Automation and pilot interactions in night and all-weather technical operations. In AGARD CP-387, Neuillly sur Seine, France 1985

[Morawski 1999] Morawski, M.: Blickbewegungsanalyse von Pilotenverhalten; Master Thesis; University of Armed Forces; Munich; 1999

[Nielsen 1993] Nielsen, J.: Usability Engineering; AP Professional, Cambridge, MA, USA, 1993

[Nygren 1991] Psychometric properties of subjective workload measurement technique: Implications for their use in the assessment of perceive mental workload; In: Human Factors, Vol. 33(1), 1991

[Onken 1998] Onken, R.: The two faces of automation; Workshop Cockpit, Munich 1998

[Rasmussen et.al. 1994] Rasmussen, J.; Pejtersen A. M.; Goodstein L.P.: Cognitive Systems Engineering; Wiley, New York 1994

[Sarter & Woods 1994] Sarter, N.; Woods, D.D.: Pilot Interaction With Cockpit Automation II: An Experimental Study of Pilots‘ Model and Awareness of the Flight Management System; Int. Journal of Aviaiton Psychology, 4(1), 1-28, 1994

[Schulte 1996] Schulte, A.: Visuelle Routinen von Piloten; Ph.D. thesis, VDI Reihe 12 Nr. 278; Düsseldorf, Germany 1996

[Schulte & Stütz 1998] Schulte, A.; Stütz, P.: Evaluation of the Cockpit Assistant Military Aircraft CAMA; NATO System Concepts and Integration Panel, Ottawa, Canada, 1998

[Stark & Choi 1996] Stark, L.W.; Choi Yum S.: Experimental Metaphysics: The Scanpath as an epistemological Mechanism; in Ed. Hangemeister, W.H.; Stiehl, H.S.; Freska, C.: Visual Attention and Cognition; Elsevier North Holland; 1996

[Verwey 1990] Adaptable driver-car interfacing and mental workload: A review of the literature; TNO Rep. No IZF1990 B-3, Soesterberg, NL, 1990

[Walsdorf & Onken 1998] Walsdorf A., Onken R..: Intelligent Crew Assistant for Military Transport Aircraft; NATO System Concepts and Integration Panel, Ottawa, Canada, 1998

[Wickens 1992] Wickens, C. D.: Engineering Psychology and Human Performance; Harper Collins 1992